

行動遺伝学における高次積率を利用した統計モデル

データ科学研究系 調査解析グループ
助教 尾崎 幸謙

1 行動遺伝学とは

行動遺伝学 (Neale and Carldon, 1992) とは、主として双生児データを利用することで、個人差に対する遺伝と環境の影響を推定する学問である。これまでに、心理学・教育学・社会学・疫学など、個人を対象とする学問で扱われる様々な表現型 (変数) に関して研究がなされてきた。推定のロジックは次の通りである。双生児には 1 卵性双生児 (Monozygotic Twins; MZ) と 2 卵性双生児 (Dizygotic Twins; DZ) があり、1 卵性双生児は遺伝子を 100% 共有しているが、2 卵性双生児は平均的に 50% 共有している。また、双生児は基本的には同じ両親のもとで、同じような環境の下で育つ。しかし、双生児の各々が独自の経験をし (異なる友人を持つ、事故に遭うなど)、それが個人の成長・形成に影響を与える可能性もある。行動遺伝学では、これらを遺伝・共有環境・非共有環境の影響と呼んでいる。遺伝・共有環境・非共有環境は統計学的には因子分析における因子でありそれぞれ、A,C,E と表現する。A,C,E の 3 因子は独立とする。またそれらの表現型に対する影響を a,c,e と表す。各因子の分散を 1 として、1 卵性双生児と 2 卵性双生児の表現型 x に対する共分散行列をパラメタ a,c,e で構造化すると、

$$(1.1) \quad \Sigma_{MZ} = \begin{bmatrix} E[x_{1MZ}^2] & E[x_{1MZ}x_{2MZ}] \\ E[x_{1MZ}x_{2MZ}] & E[x_{2MZ}^2] \end{bmatrix} = \begin{bmatrix} a^2 + c^2 + e^2 & a^2 + c^2 \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

$$(1.2) \quad \Sigma_{DZ} = \begin{bmatrix} E[x_{1DZ}^2] & E[x_{1DZ}x_{2DZ}] \\ E[x_{1DZ}x_{2DZ}] & E[x_{2DZ}^2] \end{bmatrix} = \begin{bmatrix} a^2 + c^2 + e^2 & 0.5a^2 + c^2 \\ 0.5a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}$$

となる。ここで、 x_1, x_2 はそれぞれ、双生児の 1 人目と 2 人目に関する表現型 x を表す。添え字の MZ は 1 卵性、DZ は 2 卵性を表す。分散が $a^2 + c^2 + e^2$ となっていることから分かるように、行動遺伝学では個人差を 3 つの要因に分解している。これを ACE モデルと呼ぶ。パラメタは、以下の目的関数を最小化することで求められる。

$$(1.3) \quad f = n_{MZ} \text{vec}(S_{MZ} - \Sigma_{MZ})' W_{MZ}^{-1} \text{vec}(S_{MZ} - \Sigma_{MZ}) \\ + n_{DZ} \text{vec}(S_{DZ} - \Sigma_{DZ})' W_{DZ}^{-1} \text{vec}(S_{DZ} - \Sigma_{DZ})$$

ここで、 n はペア数、 S は標本共分散行列、 W は重みである。パラメタ推定は構造方程式モデリング (Structural Equation Modeling; SEM) を用いて最尤推定によって行われる。また、ACE モデルは、共分散行列のユニークな要素の数とパラメタ数が同じ丁度識別のモデルである。

2 個人差を説明する他の要因

行動遺伝学では ACE モデルの使用頻度が非常に高いが、ADE モデルと呼ばれるモデルもしばしば使用される。D は非相加的遺伝であり、遺伝どうしの交互作用を表す。一方、A は遺伝の中でも相加的遺伝を表し、表現型に影響を与える個々の遺伝子の総和である。したがって、行動遺伝

学では、表現型に対して独立な4つの要因の影響が本来仮定されているが、ACEモデル(そしてADEモデルも)が丁度識別であるため、同一家庭で育った双生児データでACDEモデルを識別することは不可能であった。ACDEモデルを識別するためには、同一家庭で育った双生児のデータだけでなく、例えば、出生直後に双生児のうちの1人が養子に出されるなどして別々の環境で育った双生児のデータを利用する必要があった。しかし、養子に出されるケースは非常に少ないため、これらのデータを収集することは極めて困難である。

また、A,C,D,Eの間はそれぞれ独立である仮定がモデルに課されているが、実際にはそうではないこともある。例えば、遺伝的に言語能力に優れた家系に生まれた場合には、家庭内に多くの書物があるなど、言語能力を伸ばすために適した環境が与えられている可能性がある。これは、遺伝環境相関と呼ばれ、A(あるいはD)とCの間の相関に現れる。したがって、本来はACEモデルにおいてAとCの間に相関を仮定すべきであるが、上述したようにACEモデルは丁度識別であるため、相関を仮定することはできない。また、異なる遺伝子座にある3つ以上の遺伝子間の交互作用を高次のエピスタシスと呼ぶが、これが存在することを仮定したモデル開発はなされていない。その理由は、高次のエピスタシスは2卵性双生児のD因子間相関に現れるが、ADEモデルも丁度識別モデルであるため、相関は求められないからである。遺伝環境交互作用と呼ばれる現象もある。これは、同じ遺伝子を有していたとしても、その発現が環境によって異なるという現象である(これを記述するモデルは開発されている)。このように、個人差を説明する要因は様々あるが、それらを(同時に)モデルに組み込むことは、モデル識別の問題から不可能である。また、養子データはこれらの問題のいくつかを解決できるが、収集は非常に困難である。

3 高次積率の利用による解決と今後の課題

これらのモデルを識別するために、高次積率を利用したモデルを構築した。高次積率を利用したSEMは、Shimizu and Kano (2008)で提案されている。発表者は、2次と3次の積率を利用すれば、双生児データのみからACDEモデルを識別することが可能であることを示した。さらに、シミュレーション研究によって、既存のACEあるいはADEモデルよりもバイアスの小さな推定量を与えることが示された。ACEあるいはADEモデルは共有環境Cの影響をほぼ必ず過小評価することが知られている。実際、行動遺伝学の研究から大きな共有環境の影響が推定されることは稀である。しかし、推定された共有環境の影響の小ささの幾分かはおそらくモデルのmisspecificationに原因がある。ACDEモデルによって、Cの影響の過小評価が修正されたことは、個人差の原因論の見直しにつながる可能性がある。

行動遺伝学の分野で近年注目されているのが、上に挙げた遺伝環境交互作用モデルである。このモデルを使用すれば、心理的・身体的疾病に罹るリスクの高い遺伝子を有していたとしても、その遺伝子の発現を抑える調整変数としての環境を探索することが可能である。これまではACEあるいはADEモデルをもとにして遺伝環境交互作用モデルが構築されていたが、今後ACDEモデルをベースにすることで、より適切な調整変数を探索することが可能になると考えられる。また、ACDEモデルのみならず、他のモデルについても高次積率を利用したモデル開発を行っていきたい。

参考文献

- Neale, M. C., and Carldon, (1992). *Methodology for genetic studies of twins and families*, Kluwer Academic Publishers, Dordrecht.
- Shimizu, S., and Kano, Y. (2008). Use of non-normality in structural equation modeling: Application to direction of causation, *Journal of Statistical Planning and Inference*, **138**, 3483-3491.